

# Open Source Parallel Image Analysis and Machine Learning Pipeline, Phase I

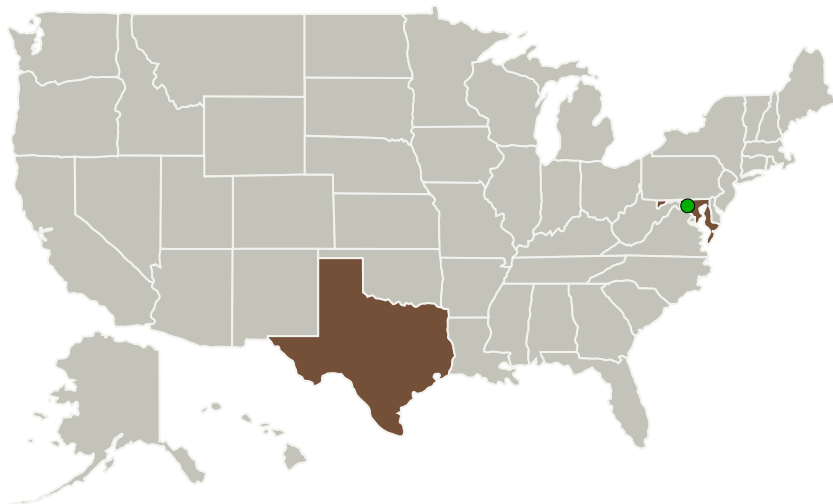
Completed Technology Project (2016 - 2016)



## Project Introduction

Continuum Analytics proposes a Python-based open-source data analysis machine learning pipeline toolkit for satellite data processing, weather and climate data processing, and machine learning and prediction with optional proprietary cluster management tools for streamlined deployment for cloud providers and on-premises clusters. The innovative software will empower scientists and analysts to readily and seamlessly construct and test workflows that transparently and scalably perform calculations across cluster nodes for data-driven discovery. The simple API for homogenous processing of images, mosaics and tiles further improves ease of use for rapid testing and prototyping of analyses paradigms for multiple extremely large data sets. Today, NASA researchers must create, debug, and tune custom workflows for each analysis. Creation and modification of custom workflows is fragile, non-portable, and consumes time that could be better spent on advancing scientific discovery. The Phase I work plan will demonstrate that it is feasible to easily create and compose data manipulations and analytics from a variety of sources with a portable, reproducible, extensible process that can be deployed on a wide variety of systems and software. This is a major improvement over the current state-of-the-art because of reduced workflow creation time, portability of deployment and use, extensibility, and robustness.

## Primary U.S. Work Locations and Key Partners



Open Source Parallel Image Analysis and Machine Learning Pipeline, Phase I

## Table of Contents

Project Introduction	1
Primary U.S. Work Locations and Key Partners	1
Project Transitions	2
Images	2
Organizational Responsibility	2
Project Management	2
Technology Maturity (TRL)	2
Technology Areas	3
Target Destinations	3

## Open Source Parallel Image Analysis and Machine Learning Pipeline, Phase I

Completed Technology Project (2016 - 2016)



Organizations Performing Work	Role	Type	Location
Continuum Analytics, Inc.	Lead Organization	Industry	Austin, Texas
● Goddard Space Flight Center(GSFC)	Supporting Organization	NASA Center	Greenbelt, Maryland

Primary U.S. Work Locations	
Maryland	Texas

## Project Transitions

▶ **June 2016:** Project Start

✓ **December 2016:** Closed out

## Closeout Documentation:

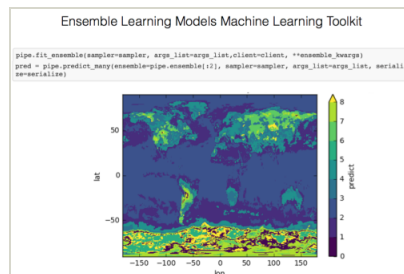
- Final Summary Chart(<https://techport.nasa.gov/file/139646>)

## Images



## Briefing Chart Image

Open Source Parallel Image Analysis and Machine Learning Pipeline, Phase I  
(<https://techport.nasa.gov/image/133931>)



## Final Summary Chart Image

Open Source Parallel Image Analysis and Machine Learning Pipeline, Phase I Project Image  
(<https://techport.nasa.gov/image/129323>)

## Organizational Responsibility

## Responsible Mission Directorate:

Space Technology Mission Directorate (STMD)

## Lead Organization:

Continuum Analytics, Inc.

## Responsible Program:

Small Business Innovation Research/Small Business Tech Transfer

## Project Management

## Program Director:

Jason L Kessler

## Program Manager:

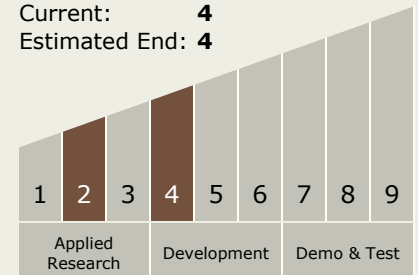
Carlos Torrez

## Principal Investigator:

Peter Steinberg

## Technology Maturity (TRL)

Start: 2  
Current: 4  
Estimated End: 4



# Open Source Parallel Image Analysis and Machine Learning Pipeline, Phase I

Completed Technology Project (2016 - 2016)



## Technology Areas

### Primary:

- TX11 Software, Modeling, Simulation, and Information Processing
  - └ TX11.4 Information Processing
    - └ TX11.4.2 Intelligent Data Understanding

## Target Destinations

The Sun, Earth, The Moon, Mars, Others Inside the Solar System, Outside the Solar System